Nonlinear time series models and their extremes, with hydrological applications

Ph.D. Thesis Summary

Péter Elek

Supervisor: András Zempléni
Associate Professor, CSc

Doctoral School of Mathematics
Director: Miklós Laczkovich

Doctoral Programme of Applied Mathematics
Director: György Michaletzky

Eötvös Loránd University
Faculty of Sciences
Department of Probability Theory and Statistics

Budapest, May 2009

1 Introduction

By combining the methods of time series analysis and extreme value theory I investigate probabilistic properties, extremal behaviour and parameter estimation of certain nonlinear time series models in this dissertation. My theoretical contributions were initially motivated by an applied hydrological project that – somewhat unusually in the hydrological literature – aimed to build models to capture both the times series dynamics and the extremal behaviour of water discharge data sets of Danube and Tisza. The coupling of the two theories may lead to a more precise description of time series extremes than their separate application would do. Purely time series models concentrate on the "typical" behaviour of the series and thus give too little weight to large observations, while purely extreme value models tend to neglect the information in the dynamics of the process, or make too general assumptions on it. The two methodologies are combined quite often in mathematical finance, actuarial studies, telecommunications but less usually in hydrological models.

The thesis summary follows the structure of the dissertation. I review the necessary preliminaries in section 2 and present my empirical findings on the properties of water discharge series in section 3. I give my theoretical and applied results on conditionally heteroscedastic models in section 4 and on Markov-switching models in section 5. Finally, section 6 reveals the relationship between the two model families and outlines directions of future research.

The results of section 4 are joint work with László Márkus, while those of section 5 are joint work with my supervisor András Zempléni.

2 Preliminaries

Although the dissertation uses preliminaries from both time series analysis and extreme value theory (EVT), I only review the basic concepts and scope of the latter field in the thesis summary. Time series analysis is a much more traditional research area hence its classical results can be found in numerous monographs.

Basically, EVT deals with two types of research problems.[1] First, the tail of a distribution (i.e. the rate of decay of the $\bar{F}(u) = 1 - F(u)$ survival function as $u \to \infty$) can be examined, and second, the clustering of high observations in a stationary series (i.e. to what extent they form groups) can also be analysed.

As far as the tail of a distribution is concerned, the theorem of Balkema-de Haan-Pickands is a basic result in EVT (Balkema and de Haan, 1974; Pickands, 1975). For any random variable X satisfying some general conditions there exists a measurable function $a(u)$ such that the distribution of normalised threshold exceedances tends to the generalised Pareto distribution (GPD) with shape parameter ξ as u tends to the upper end point of the support of the distribution:

$$((X-u)/a(u)) \mid (X>u) \to_d GPD_\xi.$$

[1] The book of Embrechts et al. (1997) is a basic monograph of EVT.

The shape parameter of the GPD obtained in the limit greatly determines the extremal properties of the original distribution as well. If $\xi > 0$ the survival function is regularly varying with parameter $1/\xi$ (it follows that $E\left(X^+\right)^m = \infty$ for $m > 1/\xi$ values),[2] while for $\xi < 0$ the support of the distribution is bounded from above and the survival function (after a simple transformation) is regularly varying around the upper end point. The $\xi = 0$ case – when the obtained GPD is the exponential distribution – can be characterised with more difficulty. Although the distributions in this group share the common feature that all of their moments are finite one can find heavy and and also light tailed distributions among them.[3] The exponential and the normal laws are examples for the light-tailed case, while the Weibull distribution with exponent smaller than one for the heavy-tailed case.[4]

During the commonly used threshold-based estimation method high quantiles of a distribution are estimated from the parameters of a GPD fitted to exceedances above a high threshold. This is an appropriate procedure in the absence of other information but leads to highly variable or (if a relatively low reference threshold was chosen) biased quantile estimates. If additional information is available on the distribution (e.g. if we not only know that it belongs to the domain of attraction of the GPD with $\xi = 0$ but know its decay more accurately) then we can obtain more precise quantile estimates. Besides theoretical interest, this gives the motivation to study the rate of decay of the stationary distribution of certain theoretical time series models – a research area that has reached deep results e.g. in mathematical finance.

Turning to the other direction of EVT, clustering of large observations in a time series means more precisely the following. Let us examine a C_n extremal functional of a stationary process X_t :

$$C_n(u) = \sum_{t=1}^{n-m+1} g\left(X_t - u, \ldots, X_{t+m-1} - u\right),$$

where g is a $\mathbb{R}^m \to \mathbb{R}_+$ function satisfying $g(\mathbf{x}) = 0$ for all $\mathbf{x} \notin \mathbb{R}^m_+$. Let us choose the $\{u_n\}$ sequence such that $\lim_{n\to\infty} n\bar{F}\left(u_n\right) \to \tau > 0$. Then under general conditions the distribution of $C_n(u_n)$ tends to the distribution of $C_1^* + C_2^* + \cdots + C_L^*$, where L is a Poisson-distributed random variable and C_1^*, $C_2^*, \ldots$ are i.i.d. variables, independent of L as well (Smith et al., 1997). Heuristically this means that large observations in a stationary time series occur in clusters, and these clusters are asymptotically independent of each other. Thus the distribution of C_i^* contains essential information about the time-dependence of extreme observations of the process. For instance, if $m = 1$ and $g(x) = \chi_{\{x>0\}}$ then C_i^* corresponds to the size of an extremal cluster in the limit (e.g. to the duration of a large flood in the hydrological context) and if $m = 1$ and $g(x) = x^+$ then the aggregate excess during an extreme event (e.g. the flood volume in hydrology) is obtained. In this case C^* will be denoted by W^*.

The simplest measure of extremal dependence is the extremal index (θ), which is obtained – under some assumptions – from the relationship $E\left(L\right) = \theta\tau$ or as the reciprocal of the expectation of the

[2] The notations $x^+ = \max\left(x, 0\right)$ and $x^- = \max\left(-x, 0\right)$ are used.

[3] A distribution is light-tailed if there exists an $s > 0$ such that the $L_X\left(s\right) = E\exp\left(sX\right)$ moment generating function is finite. If such an s does not exist the distribution is heavy-tailed.

[4] According to the terminology of the dissertation the survival function of the Weibull distribution is $\bar{F}\left(u\right) = \exp\left(-\lambda u^d\right)$ where d is the exponent parameter. A distribution has Weibull-like tail if there exist $K_1 > 0$, K_2, $\lambda > 0$ and $d > 0$ constants such that $\bar{F}\left(u\right) \sim K_1 u^{K_2} \exp\left(-\lambda u^d\right)$ as $u \to \infty$.

limiting cluster size: $\theta = E\left(C_i^*\right)^{-1}$ if $g(x) = \chi_{\{x>0\}}$. Apart from pathological cases $0 < \theta \leq 1$ and a smaller extremal index implies a stronger extremal dependence.

Since the extremal index and the other characteristics of extremal dependence are asymptotic concepts they can be estimated from finite samples (similarly to high quantiles) only with large uncertainty. Things are made even more complicated because – contrary to the GPD in the estimation of high quantiles – the limiting cluster size and limiting aggregate excess distributions generally cannot be described with parametric families. The problem is often tackled by restricting the set of examined models, deriving theoretical extremal dependence properties for the special models and then developing methods for estimating cluster characteristics based on these results. If model specification is correct these procedures are more accurate than the methods applicable for more general series as well.

A good example for this research direction is the analysis of extremal dependence of Markov chains. Under general conditions Smith et al. (1997) have shown that a Markov chain with exponential tail behaves asymptotically (above high thresholds) as a random walk. Based on this result, an estimation and simulation method has been developed for the analysis of the extremal clusters of Markov chains.

3 Empirical features of water discharge data

In the dissertation I use daily water discharge data of three monitoring stations at river Danube (Komárom, Nagymaros, Budapest) and three at river Tisza (Tivadar, Vásárosnamény, Záhony). My first results concern the empirical properties of the series. First I fit ARMA models with seasonal components to the data and then generate synthetic water discharge series with the fitted models. Following the hydrological literature (e.g. Montanari et al., 1997), the ARMA innovations are not obtained from a parametric distribution but by the bootstrap method, i.e. by resampling the fitted innovations. My results show that the simulated processes substantially underestimate the high quantiles and do not reproduce the probability density of the observed data.

Since various heuristic estimators (correlogramm-based procedure, R/S statistic etc.) point to the presence of long memory I also estimate a fractional ARIMA model on the data. However, the fit of the high quantiles and the probability density improve only marginally. Hence, contrary to other rivers (e.g. Montanari et al., 1997), linear models do not describe accurately the behaviour of daily river flows of Danube and Tisza, which makes nonlinear modelling necessary.

I also show in the chapter that – similarly to other rivers with medium or large catchments – the data sets for Danube and Tisza tend to belong to the domain of attraction of the GPD with $\xi = 0$ (the $\xi = 0$ hypothesis cannot be rejected for most monitoring stations and thresholds). The extremal index estimates reveal that observations above high thresholds are strongly clustering over time. These empirical results steer the modeler in the choice among the various possible nonlinear model families.

4 Conditionally heteroscedastic models

First I examine ARMA-β-TARCH models in the dissertation from a theoretical point of view:

$$
\begin{aligned}
X_t &= c + \sum_{i=1}^{p} a_i (X_{t-i} - c) + \sum_{i=1}^{q} b_i \varepsilon_{t-i} && (1) \\
\varepsilon_t &= \sigma(X_{t-1}) Z_t, && (2) \\
\sigma^2(x) &= \alpha_0 + \alpha_{1+} \left((x-m)^+\right)^{2\beta} + \alpha_{1-} \left((x-m)^-\right)^{2\beta}, && (3)
\end{aligned}
$$

where I make the following assumptions:

Assumption 4.1. *Z_t is an independent identically distributed random sequence with zero mean and unit variance. The distribution of Z_t is absolutely continuous with respect to the Lebesgue-measure, and its support is the whole real line.*

Assumption 4.2. *For the characteristic polynomials*

$$\Phi(z) = 1 - \sum_{i=1}^{p} a_i z^i \neq 0 \quad \textit{and} \quad \Psi(z) = 1 + \sum_{i=1}^{q} b_i z^i \neq 0 \quad \textit{if} \quad |z| \leq 1,$$

and $\Phi(z)$ and $\Psi(z)$ have no common zeros.

Assumption 4.3. $0 < \beta < 1$.

Assumption 4.4. $\alpha_0 > 0$, $\alpha_{1+} \geq 0$ *and* $\alpha_{1-} \geq 0$.

Thus, in contrast to the usual (T)ARCH-type processes, the variance of the ε_t innovation depends on X_{t-1} and not on ε_{t-1}, and the function describing this relationship tends to infinity in a slower than quadratic rate (since $\beta < 1$).

If $\beta = 1$ the model given by (1)-(3) does not have a stationary solution for every $\alpha_{1+} \geq 0$ and $\alpha_{1-} \geq 0$, and the domain of stationarity depends not only on the variance parameters but on the coefficients of the ARMA equation, too. If the stationary solution exists its survival function is polynomially decaying even for light-tailed (e.g. normally distributed) Z_t noises, hence the distribution belongs to the domain of attraction of a GPD with $\xi > 0$ (see e.g. Embrechts et al. (1997) in a special case).

Using the drift condition for the stability of Markov chains (Meyn and Tweedie, 1993) I prove that the $0 < \beta < 1$ case is substantially different from the quadratic case:[5]

Theorem 1. *If Assumptions 4.1-4.4 hold the X_t process defined by* (1)-(3) *is geometrically ergodic and has a unique stationary distribution. If, moreover, $E(|Z_t|^r) < \infty$ for an $r \geq 2$ real, then $E(|X_t|^r) < \infty$ under the stationary distribution.*

[5] Unless stated otherwise, all probability statements in the sequel hold under the stationary distribution.

It follows that if all moments of the generating noise are finite then the stationary distribution of X_t can only belong to the domain of attraction of the GPD with $\xi = 0$. In the case without ARMA-terms, if the generating noise has Weibull-like tail with exponent parameter γ then I prove a more accurate statement: the tail of the stationary distribution can be approximated by Weibull-like distributions with parameter $\gamma(1-\beta)$.

Assumption 4.5. *Z_t is an i.i.d. sequence and there exist $u_0 > 0$, $\gamma > 0$, $K_1 > 0$ and K_2 such that its probability density satisfies*

$$f_{Z_t}(u) = K_1 |u|^{K_2} \exp\left(-\kappa |u|^{\gamma}\right)$$

for every $|u| > u_0$.

Theorem 2. *Assume $a_i = 0$ and $b_i = 0$ $(i = 1, \ldots, \max(p,q))$, Assumption 4.5, $\alpha_0 > 0$, $\alpha_{1+} > 0$, $\alpha_{1-} > 0$ and $0 < \beta < 1$. Then, using the notation $\alpha_1^{\max} = \max(\alpha_{1+}, \alpha_{1-})$,*

$$\exp\left(-\frac{\left(\alpha_1^{\max}\right)^{-\gamma/2}\kappa\gamma\beta^{-\frac{\beta}{1-\beta}}}{2}u^{\gamma(1-\beta)} + O\left(u^{\gamma(1-\beta)/2}\right)\right) \leq \bar{F}_{X_t}(u)$$

$$\leq \exp\left(-\frac{\left(\alpha_1^{\max} + \alpha_0\right)^{-\gamma/2}\kappa\gamma\beta^{-\frac{\beta}{1-\beta}}}{2}u^{\gamma(1-\beta)} + O\left(u^{\gamma(1-\beta)/2}\right)\right).$$

Proposition 3. *If the assumptions of the previous theorem hold but $\alpha_{1-} > 0$ is replaced to $\alpha_{1-} = 0$, then for every $\delta > 0$ there exists a $K > 0$ such that $\exp\left(-Ku^{(1+\delta)\gamma(1-\beta)}\right) \leq \bar{F}_{X_t}(u)$.*

Finishing the probabilistic analysis, I illustrate the conjecture that – unlike in the $\beta = 1$ case – the process has a unit extremal index thus asymptotically the exceedances do not form clusters. (This does not rule out, however, clustering above large but finite thresholds.)

Turning to parameter estimation, let $\boldsymbol{\theta} = (a_1, \ldots, a_p, b_1, \ldots, b_q)$ and $\boldsymbol{\alpha} = (\alpha_0, \alpha_{1+}, \alpha_{1-})$. Denote the true ARMA parameter vector by $\boldsymbol{\theta}^0$ and the true parameters of the variance equation by $\boldsymbol{\alpha}^0$.

For given β and m I estimate the ARMA parameters by least squares and the parameters of the variance equation by QML (quasi maximum likelihood, i.e. by maximising the likelihood function obtained under the assumption of normally distributed noise sequence). In the latter part of the estimation the $\hat{\varepsilon}_t$ innovations calculated from the ARMA fit are used. I prove the consistency and asymptotic normality of the procedure under the following assumptions.

Assumption 4.6. *There exists a $\delta > 0$ such that $\boldsymbol{\theta}^0 \in \boldsymbol{\Theta}_\delta$ where*

$$\boldsymbol{\Theta}_\delta = \{\boldsymbol{\theta} \in \mathbf{R}^{p+q} : \textit{the roots of } \Phi_{\boldsymbol{\theta}}(x) \textit{ and } \Psi_{\boldsymbol{\theta}}(x) \textit{ have moduli } \geq 1+\delta\}.$$

Moreover, $\boldsymbol{\alpha}^0 \in \mathrm{int}(\mathbf{K})$, where $\mathbf{K}$ is a compact subset of $\mathbf{R}_{++} \times \mathbf{R}_{+} \times \mathbf{R}_{+}$.

Assumption 4.7. *$E(|Z_t|^{4+2\eta}) < \infty$ holds for some $\eta > 0$.*

Theorem 4. *Under Assumptions 4.1-4.3 and 4.6 the QML estimator is consistent, i.e. $\hat{\boldsymbol{\alpha}}_n \to \boldsymbol{\alpha}^0$ a.s. If in addition Assumption 4.7 holds, the resulting estimator is asymptotically normally distributed, i.e.*

$$\sqrt{n}(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}^0) \to_d N\left(\mathbf{0}, \mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\mathbf{V}\left(\boldsymbol{\alpha}^0\right)\mathbf{H}^{-1}\left(\boldsymbol{\alpha}^0\right)\right)$$

where

$$\mathbf{V}(\boldsymbol{\alpha}) = E_\pi\left(\frac{\partial l(\varepsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\frac{\partial l(\varepsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}^T\right)$$
$$\mathbf{H}(\boldsymbol{\alpha}) = E_\pi\left(-\frac{\partial^2 l(\varepsilon_t, X_{t-1}, \boldsymbol{\alpha})}{\partial^2 \boldsymbol{\alpha}}\right).$$

The $\mathbf{H}(\boldsymbol{\alpha}^0)$ and $\mathbf{V}(\boldsymbol{\alpha}^0)$ matrices can be consistently estimated by the empirical counterparts of $\mathbf{H}(\hat{\boldsymbol{\alpha}}_n)$ and $\mathbf{V}(\hat{\boldsymbol{\alpha}}_n)$, with expectations replaced by sample averages.

Finally I fit the ARMA-β-TARCH model (with $\beta = 1/2$) to the water discharge data and obtain highly significant heteroscedasticity for all monitoring stations. Synthetic river flow series are generated from the fitted model in such a way that the noise sequences are obtained by resampling the $\{\hat{Z}_t\}$ calculated noises. The simulated time series approximate the probability densitis and high quantiles of observed water discharges much better than the simulations from the linear models do.

5 Markov-switching autoregressive models

In this chapter the Markov-switching (MS-) AR(1) model family is examined:

$$X_t = a_1 X_{t-1} + \varepsilon_{1,t} \quad \text{ha} \quad I_t = 1, \tag{4}$$
$$X_t = a_0 X_{t-1} + \varepsilon_{0,t} \quad \text{ha} \quad I_t = 0. \tag{5}$$

Here I_t is a two-state discrete time Markov chain with $p_i = P\left(I_t = 1 - i | I_{t-1} = i\right)$ transition probabilities $(i = 0, 1)$, and $\{\varepsilon_{1,t}\}$, $\{\varepsilon_{0,t}\}$ are i.i.d. sequences, independent of each other and from $\{I_t\}$ but not necessarily identically distributed with each other. Let us also assume that $|a_1| \geq |a_0|$.

The EVT literature had earlier dealt mainly with the $|a_0| < 1 < a_1$ case. Under this assumption the stationary distribution – if it exists – is heavier tailed than the generating $\varepsilon_{1,t}$ noise, moreover, the survival function is polynomially decaying under general assumptions (Saporta, 2005). The extremal index of the model is smaller than one. If, in contrast, $|a_0| \leq |a_1| < 1$ the stationary distribution can take many forms but it is certainly light-tailed with a unit extremal index in case of light-tailed $\varepsilon_{i,t}$ $(i = 0, 1)$ noises.

I analyse the case not examined previously from an extremal point of view where the first regime is a random walk and the second is a stationary autoregression (Assumption 5.1). I also assume that $\varepsilon_{1,t}$ and $\varepsilon_{0,t}$ are light-tailed.

Assumption 5.1. $a_1 = 1$ *and* $0 \leq a_0 < 1$.

Assumption 5.2. *The distribution of $\varepsilon_{1,t}$ is absolutely continuous with respect to the Lebesgue-measure and $E|\varepsilon_{1,t}| < \infty$. Moreover, there exists a $\kappa > 0$ such that $(1 - p_1) L_{\varepsilon_{1,t}}(\kappa) = 1$ and $L'_{\varepsilon_{1,t}}(\kappa) < \infty$.*

Assumption 5.3. *The distribution of $\varepsilon_{0,t}$ is absolutely continuous with respect to the Lebesgue-measure and its support is the whole real line. There exists an $s_0 > \kappa$ such that $L_{|\varepsilon_{0,t}|}(s_0) < \infty$.*

If Assumption 5.1 holds the model always has a stationary solution. Assumption 5.2 plays a crucial role in examining the extremal behaviour. Let

$$S_0 = 0, \qquad S_n = S_{n-1} + \varepsilon_n \quad (n = 1, 2, \dots)$$

be a random walk where the distribution of $\{\varepsilon_n\}$ is the same as the distribution of $\{\varepsilon_{1,t}\}$. Furthermore, let T be a Geom (p_1)-distributed random variable, independent of $\{\varepsilon_n\}$. I prove a Cramer-Lundberg-type approximation for the maximum of the random walk stopped at time $T-1$ and then show the following:

Proposition 5. *Under Assumption 5.2 there exists a $K > 0$ such that $P(S_T > u) \sim K \exp(-\kappa u)$.*

This statement, combined with the drift condition for Markov chains, eventually leads to the following theorem:

Theorem 6. *If Assumptions 5.1–5.3 hold there exists a $K > 0$ such that*

$$P(X_t > u) \sim K \exp(-\kappa u).$$

Let τ_u denote the entrance time to (u, ∞) and define the $B_u = S_{\tau_u} - u$ overshoot on the $(\tau_u < \infty)$ event. Assumption 5.2 ensures that $B_u | (\tau_u < \infty) \to_d B_\infty$ as $u \to \infty$. Define the $\{S_n^*\}$ random walk as

$$S_0^* = B_\infty, \qquad S_n^* = S_{n-1}^* + \varepsilon_n \quad (n = 1, 2, \dots),$$

where the B_∞ variable is chosen independently of the $\{\varepsilon_n\}$ sequence. With these notations I prove the following proposition on the extremal clustering behaviour of the MS-AR(1) model:

Proposition 7. *Let $g(x) = 0$ for $x < 0$ and $g(x) = o(\exp(\kappa x))$ as $x \to \infty$. Then as $n \to \infty$ $C_n(u_n)$ converges to a Poisson sum of independent random variables, distributed as C^* where*

$$C^* = \sum_{k=0}^{T-1} g(S_k^*).$$

The θ extremal index is given by

$$\theta = \int_{-\infty}^{0} \kappa \exp(\kappa x) Q(x) dx$$

where $Q(x)$ is the solution of the Wiener-Hopf-equation:

$$Q(x) = p_1 + (1 - p_1) \int_0^\infty Q(y) f_{\varepsilon_{1,t}}(x - y) dy.$$

Hence, as far as extremes are concerned, this parameter choice lies between the two previously mentioned cases ($a_1 > 1$ and $|a_1| < 1$): the stationary distribution has an exponential tail but the extremal index is smaller than one. The exponent κ can be explicitly calculated in some special cases (e.g. for normally or Gamma-distributed noise) and the extremal index in others (e.g. for nonnegative or Laplace-distributed noise).

If, for instance, $\varepsilon_{1,t} \geq 0$ a.s. the limiting cluster size distribution is geometric with p_1 parameter and the extremal index is p_1. In a more special, hydrologically important case I prove a theorem about the limiting aggregate excess distribution using Laplace's method for sums:

Theorem 8. *If Assumptions 5.1 and 5.3 hold and $\varepsilon_{1,t} \sim$ Gamma (α, λ) then there exist $K_i > 0$ $(i = 1, 2)$ constants such that*

$$K_1 \exp\left(-2^{3/2}\left(\lambda_0^{-1} - \alpha\lambda_0\right)(\lambda y)^{1/2}\right) \leq \bar{F}_{W^*}(y) \leq K_2 \exp\left(-2\left(\lambda_0^{-1} - \alpha\lambda_0\right)(\lambda y)^{1/2}\right),$$

where λ_0 is the unique real number satisfying

$$\lambda_0^{-2} - 2\alpha \log \lambda_0 + \log(1 - p_1) - \alpha(1 + \log \alpha) = 0.$$

The examined MS-AR(1) model is interesting not only on its own right but in the analysis of extremal dependence of more general processes as well. Let X_t satisfy the following conditions:

Assumption 5.4. *Let I_t be a discrete time Markov chain as above. Let X_t be a stationary process whose conditional distribution, provided that I_t is known, only depends on the value of X_{t-1} (i.e. X_t is conditionally Markov in each regime). Formally, for $A_t \subset \mathbf{R}$ Borel-sets and $j_t \in \{0, 1\}$,*

$$P\left(X_t \in A_t | I_t = j_t, X_{t-i} \in A_{t-i}, I_{t-i} = j_{t-i}, i = 1, 2, \dots\right) = P\left(X_t \in A_t | X_{t-1} \in A_{t-1}, I_t = j_t\right).$$

Moreover, for each t, conditionally on $(I_1, I_2, \dots, I_t)$, the set of random variables $(X_1, X_2, \dots, X_t)$ is independent of $(I_{t+1}, I_{t+2}, \dots)$.

Assumption 5.5. *The stationary distribution of X_t is absolutely continuous with repsect to the Lebesgue-measure and there exist $0 < a \leq 1$, $K_0 > 0$, $K_1 > 0$ constants such that $\bar{F}_1(u) \sim K_1 e^{-\kappa u}$ and $\bar{F}_0(u) \sim K_0 e^{-\kappa u/a}$.*

Let $a_1 = 1$ and $a_0 = a$, and use the following notations for $j = 0, 1$:

$$F_j^u(z) = P\left(X_t < a_j u + z | X_{t-1} = u, I_t = j\right).$$

If Assumptions 5.4 and 5.5 are satisfied and the $(X_{t-1}, X_t) \,|\, (I_t = j)$ distributions $(j = 0, 1)$ belong to the domain of attraction of a bivariate extreme value law then (under further regularity conditions) $F_j^u(z)$ can be shown to have a limit for all z as $u \to \infty$. Instead of stating the regularity conditions precisely I formulate this as an assumption.

Assumption 5.6. *The joint distributions* $(X_{t-1}, X_t) \,|\, (I_t = j)$ $(j = 0, 1)$ *are absolutely continuous with respect to the Lebesgue-measure. There exist (possibly improper) distribution functions* $F_j^*(z)$ *such that* $F_j^u(z) \to F_j^*(z)$ *as* $u \to \infty$ *uniformly on all compact intervals* $(j = 0, 1)$. *Moreover, if* $F_j^*(-\infty) = \lim_{z \to -\infty} F_j^*(z) > 0$ *for a* j, *then*

$$\lim_{M \to \infty} \limsup_{u \to \infty} \sup_{y \geq M} P\left(X_t > a_i u | X_{t-1} = u - y, I_t = i\right) = 0$$

is satisfied for $i = 0, 1$.

Let us define Y_t as an MS-AR(1) process with $a_1 = 1$, $a_0 = a$ and F_j^*-distributed $\varepsilon_{j,t}$ noises. I prove that the behaviours of X_t and Y_t do not differ substantially from each other above high thresholds:

Proposition 9. *If Assumptions 5.4-5.6 hold then for all* p, $j_t \in \{0, 1\}$ *and* y_t $(t = 1, \ldots, p)$

$$\lim_{u \to \infty} \left| P\left(X_t < \left(\prod_{i=1}^{t} a_{j_i}\right) u + y_t \,(t = 1, \ldots, p) \,| X_0 = u, I_t = j_t \,(t = 1, \ldots, p)\right) \right.$$
$$\left. - P\left(Y_t < \left(\prod_{i=1}^{t} a_{j_i}\right) u + y_t \,(t = 1, \ldots, p) \,| Y_0 = u, I_t = j_t \,(t = 1, \ldots, p)\right) \right| = 0.$$

Hence (under regularity conditions similar to those routinely applied in the statistical practice) the high-level clustering of Markov-switching, conditionally Markov processes with exponential tail can be approximated by the similar behaviour of MS-AR models. Therefore, if the aim is to estimate the extremal dependence structure of a time series then it may be more effective to fit an MS-AR model only to high-level exceedances rather than to the whole series. This way, only the conditional Markov and not the conditional AR structure is assumed for the process. In the dissertation I develop an approximate maximum likelihood procedure for the threshold-based estimation of the MS-AR model, examine the properties of the estimator by simulation and apply it to the water discharge data at Tivadar. The flood maxima and flood volumes simulated from the estimated model fit well to the corresponding characteristics of the observed series, implying that the extremes of the river flow data can be adequately described by a simple conditionally Markov model.

6 Conclusions

Among the models presented in the dissertation the ARMA-β-TARCH model generalises the linear family in a statistically motivated way, while the MS-AR process gives more "structural" insight into the behaviour of hydrological time series. Due to this reason, and also because of its better fitting theoretical extremal properties, the latter model is more appropriate for hydrological purposes. It is not surprising, however, that a purely statistical fitting procedure may easily lead to an ARCH-type process: the weak ARMA representation of MS-AR models can be shown to be conditionally heteroscedastic, moreover, the variance depends on the lagged value of the process approximately linearly in a wide range. This relationship connects the two model families.

Journal articles of the author

The dissertation is based on the following four peer-reviewed journal articles and on one yet unpublished manuscript:

- Elek, P., Márkus, L., 2004. A long-range dependent model with nonlinear innovations for simulating river flows. *Natural Hazards and Earth Systems Sciences* 4, 277-283.
- Elek, P., Márkus, L., 2008. A light-tailed conditionally heteroscedastic model with applications to river flows. *Journal of Time Series Analysis* 29, 14-36.
- Elek, P., Zempléni, A., 2008. Tail behaviour and extremes of two-state Markov-switching autoregressive processes. *Computers and Mathematics with Applications* 55, 2839-2855.
- Elek, P., Zempléni, A., 2009. Modelling extremes of time-dependent data by Markov-switching structures. *Journal of Statistical Planning and Inference* 139, 1953-1967.
- Elek, P., Márkus, L., 2009. Tail behaviour of β-TARCH processes. *Manuscript*, submitted.

The author has also published two conference proceedings and various conference abstracts in the topic of the dissertation, and coauthors further three journal articles in applied mathematics or statistics, which are more or less connected to the topic.

- Arató, M., Bozsó, D., Elek, P., Zempléni, A., 2008. Forecasting and simulating mortality tables. *Mathematical and Computer Modelling* 49, 805-813.
- Bíró, A., Elek,P., Vincze, J., 2008. Model-based sensitivity analysis of the Hungarian economy to shocks and uncertainties. *Acta Oeconomica* 58, 367-401.
- Vasas, K., Elek, P., Márkus, L., 2007. A two-state regime switching autoregressive model with application to river flow analysis. *Journal of Statistical Planning and Inference* 137, 3113-3126.

References

[1] Balkema, A. A., de Haan, L., 1974. Residual life time at great age. Ann. Probab. 2, 792-804.

[2] Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

[3] Meyn, S. P., Tweedie, R. L., 1993. Markov Chains and Stochastic Stability. Springer, London.

[4] Montanari, A., Rosso, R., Taqqu, M., 1997. Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation. Water Resour. Res. 33, 1035-44.

[5] Pickands, J., 1975. Statistical inference using extreme order statistics. Ann. Statist. 3, 119-131.

[6] Saporta, B., 2005. Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian coefficients. Stochastic Process. Appl. 115, 1954-1978.

[7] Smith, R. L., Tawn, J. A., Coles, S. G., 1997. Markov chain models for threshold exceedances. Biometrika 84, 249–268.